

5. Elementy statystyki opisowej

Podstawowe pojęcia statystyki. Sposoby prezentowania danych zebranych w wyniku obserwacji statystycznej

Statystyka jest nauką, która zajmuje się badaniem zjawisk masowych. Wyróżniamy jej dwa działy: statystykę opisową i statystykę matematyczną. Przedmiotem statystyki opisowej są zagadnienia związane ze zbieraniem, porządkowaniem, analizą i interpretacją zgromadzonych danych. Statystyka matematyczna jest działem rachunku prawdopodobieństwa i zajmuje się modelami matematycznymi, których używa się do badania zjawisk masowych.

Jednym z etapów badania statystycznego jest **obserwacja statystyczna**, którą przeprowadza się za pomocą wywiadu kwestionariuszowego, ankiety, monitoringu czy rejestracji. Badaniem statystycznym obejmuje się zwykle pewien zbiór obiektów, który nazywamy **populacją generalną** (zbiorowością generalną). Badanie obejmujące wszystkie elementy populacji nazywamy **badaniem pełnym**. Najczęściej przeprowadza się badanie częściowe obejmujące tylko pewną część populacji. Taki podzbiór populacji, który został bezpośrednio objęty badaniem statystycznym, nazywamy **próbą**, a liczbę elementów wchodzących w skład próby – **liczebnością** tej próby. W statystyce mówimy o małych próbach, jeśli liczebność próby jest nie większa niż 30, oraz o próbach dużych, jeśli liczebność próby jest większa niż 30. Próba, która podlega badaniu statystycznemu, powinna być odpowiednio dobrana. Struktura próby musi odzwierciedlać strukturę badanej populacji tak, aby istniała możliwość uogólnienia otrzymanych wyników na całą populację.

Elementy populacji generalnej, jakie podlegają obserwacji statystycznej, mają różne właściwości, które nazywamy **cechami statystycznymi**. W przypadku populacji danego miasta możemy mówić o następujących cechach: wiek, płeć, kolor oczu, wykształcenie, wzrost, posiadanie własnego mieszkania czy samochodu, zawód wykonywany czy stopień zamożności (tzn. dochód miesięczny na jednego członka rodziny). Wśród cech są takie, które możemy wyrazić za pomocą liczb (np. wzrost, waga, dochód na członka rodziny). Te cechy nazywamy **cechami mierzalnymi**. Są też takie cechy, które możemy wyrazić jedynie za pomocą słów (np. kolor oczu, płeć, wykształcenie); nazywamy je **cechami niemierzalnymi**.

W wyniku badania statystycznego otrzymujemy dane statystyczne. Dane te analizuje się i opracowuje. Następnie prezentuje się wnioski wynikające z uzyskanych danych.

Przykład 1.

Wśród 300 uczniów pewnego liceum przeprowadzono ankietę. Celem ankietujących było zdobycie informacji, którego przedmiotu z grupy matematyczno-przyrodniczych młodzież uczy się najchętniej. Każdy ankietowany wskazywał tylko jeden przedmiot. Okazało się, że na matematykę wskazało 90 osób, geografię – 75 osób, fizykę – 30 osób, chemię – 45 osób i biologię – 60 osób.

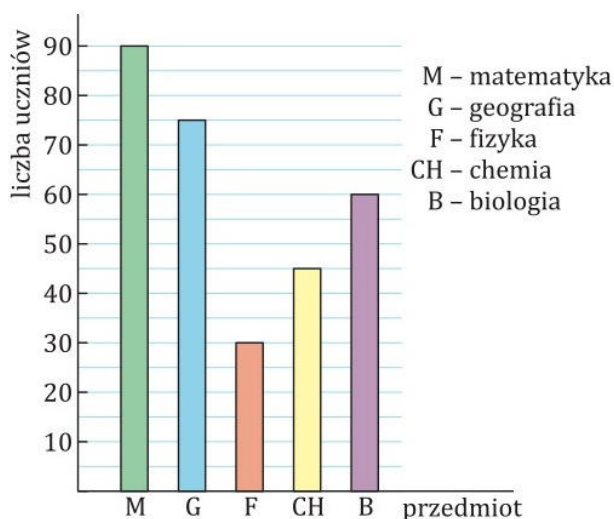
Zdobyte przez ankieterów informacje przedstawimy na różne sposoby.

- **Tabela**

Przedmiot	Liczba uczniów
Matematyka	90
Geografia	75
Fizyka	30
Chemia	45
Biologia	60

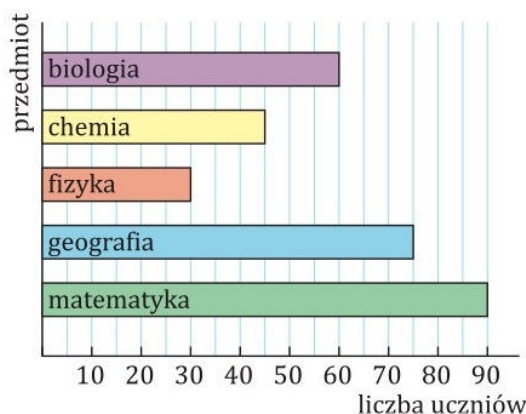
Zauważ, że tabela zawiera dwie kolumny – w pierwszej kolumnie występuje nazwa przedmiotu, w drugiej liczba uczniów, którzy wybrali dany przedmiot. Suma liczb w drugiej kolumnie wynosi 300 i jest równa liczbie ankietowanych. Przy takim przedstawieniu wyników mówimy, że tabela zawiera wyniki mierzonej cechy (w tym przypadku – nazwę przedmiotu) i jej liczebności (liczbę uczniów, którzy wskazali dany przedmiot).

- **Diagram kolumnowy**



Każda z kolumn przedstawionych na diagramie odpowiada jednemu z przedmiotów szkolnych podlegających badaniu. Legenda zamieszczona obok diagramu informuje nas, która kolumna odpowiada jakiemu przedmiotowi. Wysokość kolumny odpowiada liczbie głosów oddanych na dany przedmiot.

• Diagram słupkowy

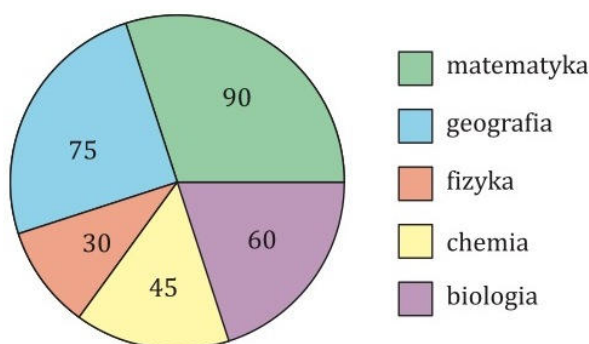


Oś pozioma informuje o liczbie głosów oddanych na dany przedmiot. Poszczególne słupki (poziome pasy) odpowiadają badanym przedmiotom. Im dłuższy słupek, tym więcej głosów oddanych na dany przedmiot.

• Diagram kołowy

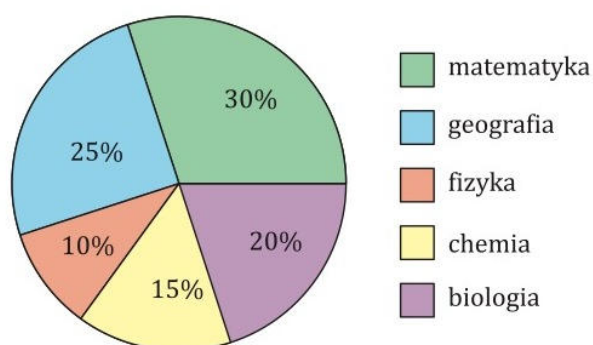
Aby sporządzić diagram kołowy, należy wykonać odpowiednie obliczenia. Interesuje nas zależność między liczbą głosów oddanych na dany przedmiot a częścią koła, która odpowiada tej liczbie głosów. Obliczenia wykonujemy w następujący sposób:

Na matematykę oddano 90 głosów, co stanowi $\frac{90}{300}$ wszystkich oddanych głosów, czyli 0,3. Zatem liczbie głosów oddanych na matematykę odpowiada kąt środkowy o mierze $0,3 \cdot 360^\circ = 108^\circ$. Podobnie wykonujemy pozostałe rachunki. Liczbie głosów oddanych na geografię odpowiada kąt środkowy o mierze 90° , na fizykę – kąt środkowy o mierze 36° , na chemię – kąt środkowy o mierze 54° , na biologię – kąt środkowy o mierze 72° . Poniżej prezentujemy zebrane dane statystyczne w postaci diagramu kołowego.



Zdarza się, że dane uzyskane w wyniku badania statystycznego przedstawia się w postaci diagramu procentowego. Diagram kołowy procentowy przedstawia zebrane dane wyrażone w procentach. Informacje otrzymane przez ankietowanych wskazują, że na matematykę głosowało 30% uczniów, na geografię – 25% uczniów, na fizykę – 10% uczniów, na chemię – 15% uczniów, a na biologię – 20% uczniów.

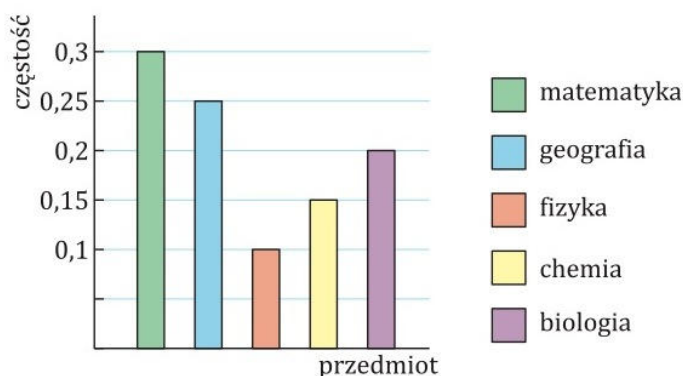
Oto diagram kołowy procentowy zebranych danych.



• Diagram częstości względnych

Opracowując wyniki ankiety, możemy wziąć pod uwagę tzw. częstości względne. Częstość względna to ułamek właściwy równy stosunkowi liczby wyników przypisanych badanej cesze (w tym przypadku liczby głosów oddanych na dany przedmiot) do ogólnej liczby wyników. W przypadku ankiety dotyczącej przedmiotów szkolnych częstość względna występowania matematyki wśród oddanych głosów wynosiła 0,3; częstość względna występowania geografii – 0,25; częstość względna występowania fizyki – 0,1; częstość względna występowania chemii – 0,15; częstość względna występowania biologii – 0,2.

Poniżej prezentujemy diagram częstości względnych poszczególnych przedmiotów.



Sprawdź, czy rozumiesz

1. Wśród pięćdziesięciu losowo wybranych osób przeprowadzono sondę uliczną, w której zapytano: „Ile razy w ciągu minionego półrocza był(a) Pan(i) na spektaklu w teatrze?”. Otrzymano następujące wyniki: 0 razy – odpowiedziało 40% badanych, 1 raz – 30% osób, 2 razy – 20% osób, 3 razy – 10% osób. Przedstaw zebrane dane w postaci tabeli liczebności, diagramu kołowego oraz diagramu słupkowego.